Hysteresis effects of changing the parameters of noncooperative games

David H. Wolpert, Michael Harré, Eckehard Olbrich, Nils Bertschinger, and Jürgen Jost

Phys. Rev. E **85**, 036102 — Published 6 March 2012

DOI: 10.1103/PhysRevE.85.036102

Hysteresis effects of changing parameters of noncooperative games

David H. Wolpert,[1,*] Michael Harré,[2] Eckehard Olbrich,[3] Nils Bertschinger,[3] and Jürgen Jost[3,†]

[1]*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA*
[2]*Centre for the Mind, The University of Sydney, Australia*[‡]
[3]*Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, D-04103 Leipzig, Germany*

We adapt the method used by Jaynes to derive the equilibria of statistical physics to instead derive equilibria of bounded rational game theory. We analyze the dependence of these equilibria on the parameters of the underlying game, focusing on hysteresis effects. In particular, we show that by gradually imposing individual-specific tax rates on the players of the game, and then gradually removing those taxes, the players move from a poor equilibrium to one that is better for all of them.

INTRODUCTION

One of the most succinct derivations of statistical physics is based on the Maximum Entropy (Maxent) principle of information theory [1–3]. In this paper we apply Maxent to the problem of predicting the joint behavior of interacting humans rather than the problem of predicting the joint behavior of interacting particles. This provides a connection between statistical physics, information theory, and game theory.

Maxent concerns the problem of how best to predict the probability distribution p over a system's states based on limited prior knowledge concerning p. It says we should make that prediction using a version of Occam's razor: Choose the p that assumes nothing beyond the prior knowledge. Maxent formalizes this version of Occam's razor as meaning we should choose the p that has maximum entropy among all p consistent with that prior knowledge. To solve for that p we must extremize a Maxent Lagrangian.

In the context of statistical physics, our prior knowledge might be the Hamiltonian of the system and the value of the system's expected energy. In this case the Maxent Lagrangian is the system's free energy. So the Maxent principle says that for that prior knowledge, we should predict the p that minimizes the free energy of the system, i.e., predict p is the canonical ensemble [3, 4]. If our prior knowledge also specifies the expected number of particles in the system, p is defined over a different space, and Maxent now says we should predict that p is the grand canonical ensemble.

Noncooperative game theory [5–8] is concerned with how to predict the behavior of a set of human "players", based on knowing each player's utility function. This is a starting point for the field of microeconomics. More generally, it is central to the foundations of many formalizations of socio-economic systems. It also has proven central to many analyses of natural selection, in the guise of evolutionary game theory [9, 10].

The idea of applying tools from statistical physics to microeconomic extends back to the 19th century. While some more recent work on this issue has come from the economics community [11], most of the recent work has originated in the physics community, under the label of "econophysics". Much of econophysics is concerned with socio-economic systems at a "coarse grained level", for instance using the tools of statistical physics to analyze stylized facts of empirical distributions such as the returns in financial markets [12]. In addition to such analysis of statistical regularities in large financial and economic data sets, econophysics also includes models of interactions of possibly heterogeneous agents.

Often (but not always), this work either does not consider the foundations of the behavior, or tries to provide explanations using non-game theoretic, mechanistic models. Some examples are multi-agent models with agents choosing their behavior according to some plausible rules [13]. Other examples are models from physics such as Ising or Potts models that are used as abstract models for collective phenomena, e.g. in option dynamics [14] or herd behavior in stock markets [15]. Typically this work tries to deduce or predict statistical regularities at the coarse-grained macrolevel as emerging from the local interactions of many such agents and to understand the mechanisms of aggregation. Among many references, a sample of recent achievements is [16] (minority game models), [17] (global effects of local behavior), [18] (heterogeneous agents and phase transitions). See [19] for a nice overview.

In this paper we start by showing how the techniques of statistical physics can also be applied to analyze socio-economic systems at the fine-grained level of game theory, as well at the coarse-grained level usually considered. Elaborating the analysis of [20], we begin by showing how to use the Maxent principle to derive a modification to the Nash equilibrium concept of game theory [5–8].

In this application of Maxent to game theory, we have a separate piece of prior knowledge for each player of the game, concerning the expected value of the utility of that player. In comparison, when using Maxent to derive the canonical ensemble, we have a single piece of prior knowledge, which also concerns an expected value (of the energy of the full system). Due to the formal similarity of these two types of prior knowledge, the modified version of the Nash equilibrium that we derive is similar to the canonical ensemble. However rather than a single Boltzmann distribution, involving a single Hamiltonian and a single temperature, our modification to the Nash equilibrium has a separate Boltzmann distribution for each player. Each player's Boltzmann distribution involves that player's utility function, and a "temperature" unique to that player. (This temperature is a quantification of the associated player's bounded rationality.) We also briefly discuss the application of Maxent to game theory when the number of players of various types in unknown. This results in each player type following a grand canonical ensemble rather than a canonical ensemble.

After deriving our modified Nash equilibrium that is related to the canonical ensemble, we analyze its dependence on the parameters of the underlying game, focusing on bifurcation behavior and hysteresis effects. In particular, we show that the changes of the "temperature" can be interpreted as changes of a "tax rate" and that by "adiabatically slowly" increasing individual-specific tax rates on the players of the game, and then gradually removing those taxes, the joint behavior of the players moves from a poor equilibrium to a Pareto-superior one. In fact, this can be done in such a way that the players agree to each infinitesimal change in tax rates, since each such change increases their expected utility.

Next we introduce three toy models of how a society may collectively decide on each such infinitesimal change in tax rates. One of these is a myopic model of "socialism", in which each change is made to maximize the immediate gain in the sum of the player expected utilities. Another is a myopic model of a "market", in which the players use unstructured bargaining [6] to decide on the infinitesimal changes in their tax rates. The final model is a myopic version of "anarchy", in which each player changes their own tax rate, assuming the others do not change theirs.

We end up by comparing these three ways of running a society in terms of the associated discounted sum of total utilities along the path of tax rates. We find that the anarchy model always does worse than the other two models. However the market model outperforms the socialism model for a low enough discounting rate, whereas socialism does better near-term, i.e., with a large enough discounting rate.

BACKGROUND

The maximum entropy principle

Shannon [21] was the first person to realize that based on a simple set of axioms, there is a unique real-valued quantification of the amount of syntactic information in a distribution $p(y)$. This quantification is the Shannon entropy of that distribution, $S(p) = -\sum_u p(y) ln[\frac{p(y)}{\mu(y)}]$. It measures the amount of uncertainty in p concerning an outcome y generated by sampling p. As such, it can be seen as the amount of information that can be gained from observing an outcome sampled according to p.[1]

As an example, the distribution with maximal entropy, i.e., highest uncertainty, is the one that doesn't distinguish at all between the various y; the uniform distribution. Conversely, the most precise distribution is the one that specifies a single possible y. For this distribution, we cannot gain any further information by observing an outcome, because we know already which outcome — y — will appear. Note that for a product distribution, entropy is additive, i.e., $S(\prod_i p_i(y_i)) = \sum_i S(p_i)$.

Say we are given some incomplete prior knowledge about a distribution $P(y)$. How should one estimate $P(y)$ based on that prior knowledge? Shannon's result (as interpreted by Jaynes [3]) tells us how to do that in the most conservative way: do not put anything else into your estimate of $P(y)$ beyond what is already contained in the prior knowledge about $P(y)$. Information about what is uncertain, i.e., not yet known, should be gained from observations and not assumed prior to them. This approach is called the maximum entropy principle (**Maxent**).

As an example, the prior knowledge concerning $P(y)$ may be in the form of one or more constraints on expected values of functions under P. Used this way, Maxent has proven extremely accurate in domains ranging from signal processing to image processing to supervised learning [2]. Famously, it was also used by Jaynes to derive statistical physics [4]; the prior knowledge constraints in that domain concern quantities like the expected energy of a system or its expected number of particles of various types.

Noncooperative game theory

In a finite, strategic form noncooperative game, one has a set of N **players**. Each player i has its own set of allowed **pure strategies**, X_i. A **mixed strategy** is a distribution $q_i(x_i)$ over player i's $|X_i|$ possible pure strategies. We write the joint space of all players' pure strategies as X. The joint distribution over X is given by sampling each player's mixed strategy independently: $q(x) = \prod_i q_i(x_i)$. As shorthand, we will use the minus symbol to indicate the set of all players with one removed, e.g., $q_{-i}(x_{-i}) \equiv \prod_{j \neq i} q_j(x_j)$. We call a joint pure (mixed) strategy choice of all the players a pure (mixed) **strategy profile**.

Each player i has a **utility function** $u_i : X \to \mathbb{R}$. So given mixed strategies of all the players, the expected utility of player i is $\mathbb{E}(u_i) = \sum_x \prod_j q_j(x_j) u_i(x)$. Much of noncooperative game theory is concerned with **equilibrium concepts** specifying what joint-strategy one should expect to result from a particular game. In particular, in a **Nash equilibrium** every player adopts the mixed strategy that maximizes its expected utility, given the mixed strategies of the other players. More formally, $\forall i$,

$$q_i = \text{argmax}_{q'_i} \Big[\sum_x p'_i(x_i) q_{-i}(x_{-i}) u_i(x) \Big].$$

In general, this set of coupled equations has multiple solutions.

A well-recognized problem with using the Nash equilibrium concept as a way to make predictions concerning the real world is its assumption of (**common knowledge** of) **full rationality**. This is the assumption that every player i can both

---

[1] In this equation μ is an *a priori* measure over y, allowing the argument of the logarithm to be unitless. It is often interpreted as a prior probability distribution. Unless explicitly stated otherwise, in this paper we will always assume it is uniform, and not write it explicitly. See [1, 3, 4].

calculate what the strategies $q_{j\neq i}$ will be and then calculate its associated optimal distribution. This assumption has been found to be (sometimes badly) violated in many experimental settings [22, 23]. Below we provide a modified version of the Nash equilibrium that accommodates bounded rationality.

A central feature of all noncooperative game theory, bounded rational or otherwise, is that the players are "strategic": each player i uses her knowledge concerning the utility functions of the *other* players to predict the behavior of those other players, presuming they will do the same concerning her. Player i then uses that prediction together with her own utility function to decide how she will behave. (In contrast, "non-strategic" models are more like mean-field models, in that each player i presumes that other players ignore her.)

MAXENT AND QUANTAL RESPONSE EQUILIBRIA

Maxent noncooperative games

To predict what q the players in a given N-player game Γ will adopt, first pick one of the players, i. Consider a counter-factual situation, where i has the same move space and utility function as in Γ, but rather than have a set of $N-1$ other humans set the distribution over X_{-i}, an inanimate stochastic system sets that distribution, to some $q_{-i}(x_{-i})$. In general, due to her limited knowledge of q_{-i}, limited computational power, etc., i will choose a suboptimal q_i, i.e., $q_i \notin \text{argmax}_{p_i}[\mathbb{E}_{p_i q_{-i}}(u_i)]$.[2] To quantify this bounded rationality, in analogy to Jaynes' derivation of the canonical ensemble, presume that player i is good enough at choosing her mixed strategy q_i so that $\mathbb{E}_{q_i,q_{-i}}(u_i)$ is some (nonmaximal) value K_i for the given q_{-i}.

Writing it out explicitly, for each player i the Maxent Lagrangian associated with this constraint is

$$\mathscr{L}_i(q_i) = S(q_i) \; + \; \beta_i[K_i - \sum_{x_i} q_i(x_i)\mathbb{E}(u_i \mid x_i)] + \lambda'_i[1 - \sum_{x_i} q_i(x_i)] \quad (1)$$

where the Lagrange parameters are β_i and λ'_i, q_{-i} is implicit, and as usual $q(x) = \prod_i q_i(x_i)$. The normalized q_i that maximizes the Lagrangian in Eq. (1) is

$$q_i(x_i) = \frac{e^{\beta_i \mathbb{E}(u_i|x_i)}}{\sum_{x'_i} e^{\beta_i \mathbb{E}(u_i|x'_i)}} \quad (2)$$

Note that as $\beta_i \to \infty$, i becomes increasingly rational, whereas as $\beta_i \to 0$, she becomes increasingly irrational; rational people are cold and irrational people are hot, using the analogy of β_i with an inverse temperature.

[2] Whereas physics systems "want to minimize" the value of their Hamiltonian, humans want to maximize the value of their utility function.

Next, recall that by the axioms of utility theory [24], *all* that player i is concerned with in choosing her mixed strategy is the resultant expected utility. Accordingly, we presume that if the best i can do is choose a particular q_i when q_{-i} is set by an inanimate system, she would also choose q_i if she faces that same distribution q_{-i} when it is set by other humans.

There is nothing in the foregoing that is particular to player i. So Maxent predicts that Eq. 2 should hold simultaneously for all N players i, for the appropriate player-specific K_i and u_i (and therefore for Lagrange parameters β_i that are player-specific). This gives a set of N coupled non-linear equations for q. Brouwer's fixed point theorem [25] guarantees that that set always has a solution, and it might have more than one.[3]

Usually in Maxent the constraints are exact equalities for the expectation values, e.g., in the derivation of the canonical ensemble. So we have formulated the constraints that way here. Note though that we get the same solution of Eq. (2) for each q_i if we change the optimization problem by using the weaker inequality constraint that $\mathbb{E}(u_i) \geq K_i$ rather than $\mathbb{E}(u_i) = K_i$.

This prediction for q is not based on a model of bounded rational human behavior derived from experimental data. It is based on desiderata concerning the prediction process *of the modeler external to the system*, not on a model of the system being predicted. Nonetheless, it is intriguing to note that maximizing Shannon entropy has a natural interpretation in terms of common models of human bounded rationality involving the cost of computation. To see this, recall that $-S(q_i)$ measures the amount of information in the distribution q_i, up to an overall additive constant. Say we equate the cost to i of computing q_i with this amount of information.[4] Then under the Maxent solution, player i minimizes the cost of computing her mixed strategy, subject to a lower bound on the value of her expected utility. (This lower bound acts as an "aspiration level" for player i.) Equivalently, she can be seen as maximizing her expected utility, subject to a bound on her computational cost. Under either interpretation, β_i quantifies i's cost of computing q_i, in units of expected utility.

Relation to earlier work

Solutions for q to the N coupled equations given by Eq. 2 are typically called (logit) **Quantal Response Equilibria** (QRE) in game theory [31–34]. They were originally derived under assumptions that the players are purely rational, but uncertain of one another's utility functions. They also arise in asymptotic analysis of several *ad hoc* models of how players

[3] An alternative Maxent approach would use it to set the entire joint distribution $q(x) = \prod_i q_i(x_i)$ at once, rather than use it to set each q_i separately and then impose self-consistency. However there are difficulties in choosing what constraints to use under this approach. See [20].

[4] Other models of the cost of computation can be found in [26–30].

learn over repeated plays of the same game [26]. In addition they have been independently suggested several times as an *a priori* reasonable way to model human players [26, 35–39]. (Some of this work has noted the relation between the logit distribution and statistical physics, e.g., [39].) The use of the logit distribution over possible moves by agents also has a long history in the Reinforcement Learning (RL) literature [40–43], and has also been shown to be is related to the replicator dynamics of evolutionary game theory as well as the QRE [44].

None of this earlier work has derived the use of a logit distribution from first principles considerations of the prediction problem.[5] Nor has any of it connected the logit distribution mixed strategies with information theory. In practice, the QRE is simply treated as a few-degree of freedom model of bounded rational play, and has been broadly and successfully used to fit experimental data concerning human behavior.

In addition none of the earlier work on the QRE has considered the shape of the QRE surface as a function of the parameters of the game. Nor has any of it considered the associated issue of how to change those parameters to move a set of players across the QRE surface. These are the topics of the next two sections. (Perhaps the closest results in the literature can be found in [33] and [46].)

Also, recall that the Maxent prediction that each player's mixed strategy is a logit distribution is not based on any model of human behavior. It arises from axioms concerning the prediction process *of a scientist external to the system*. This also contrasts with the earlier work, where it arises as part of a model of the system being predicted.

Finally, there has been other earlier work that is related to the analysis of this paper in that it involves path dependency of the effects of changes to parameters of an economy [47], in some instances considering "adiabatically slow" changes. Some of this earlier work explicitly considered bifurcation surfaces. In particular some of it focuses on catastrophes as paths discontinuously jump from one fold to another. (Some of that work has been been criticized for claiming to explain too many empirical phenomena; see [48] for a discussion.)

Most of this earlier work on path dependency has not involved game theoretic models, but rather has been more "coarse-grained", involving non-strategic players interacting in purely macroeconomic models [49–55]. In particular, conventional catastrophe theory is based on bifurcations surfaces involving a single potential function, whereas the work here is based on bifurcations that inherently involve multiple "potential functions" under the guise of the players' utility functions. In addition, almost all of this earlier work has assumed fully rational players, despite the huge volume of laboratory and field experimental data [22] establishing that real humans are often very non-rational.[6] In particular, none of this earlier work has explicitly considered a QRE model of players, as we do here. Note though that subsequent to a posting of an early version of this paper, some experimental work was done that validates some of the "stylized" character of the predictions that we make here for a QRE model [57].

Finally, there is some work in traditional game theory that models games with variable numbers of players. However the kinds of scenarios considered in such work differ substantially from the ones considered above in the derivation of the grand canonical QRE [58].

Beyond game theory's canonical ensemble

Whenever one's information concerning a distribution q over the states of a system can be expressed as constraints on q, it is straightforward to use Maxent to estimate q. Such constraints do not have to concern expected utility and/or computational effort of the players. As a result the Maxent approach is broadly applicable to game theory, just as it is broadly applicable to many other fields, ranging from signal processing to phyogenetic tree reconstruction to text analysis). This allows us to uncover many formal connections between game theory and statistical physics.

A very simplified example can illustrate those connections. Consider a situation where there are F total firms in a particular industry, each with a total of T possible employee types (e.g., salesmen, managers, production line workers, etc.). In general, each firm will have to decide how many employees of each type to have. Let n_i^k be the number of employees of type k of firm i, where n is the entire matrix of all such numbers. Say that for each firm i we know N_i, the expected total number of employees that firm i has, i.e., we know the values of $\mathbb{E}(\sum_k n_i^k)$ for all firms i. The utility (e.g., profit) of each firm i will depend on a huge number of variables of course. Rather than presume that we know that full dependence, say we only know the expected utility of firm i conditioned on any value of the matrix n. Write that conditional expected utility as $u_i(n)$. Finally, say we also know the unconditioned expected utility for each firm i, K_i.

There are numerous ways we might ascertain these expectation values, e.g., using observables like historical data, industry surveys, instrumental variables, etc. (Such observables play the same role here as the temperature reading on a thermometer plays in Jaynes' derivation of the canonical ensemble; indirect estimators of an expectation value that cannot be directly observed.) Presume though that we know nothing

[5] In particular the use of a logit distribution in the QRE literature is justified by appealing to the the choice theory literature [45], where it arises if we assume double-exponential noise is added to a player's perceived utility values. However that double-exponential assumption is never axiomatically justified in the choice theory literature; it is adopted for the calculational convenience of its resulting in the logit distribution over player choices.

[6] One exception is [26], which considers Hopf bifurcations under an ad hoc model for player behavior that involves Shannon entropy. Another is [56], which relaxes the full rationality assumption in a way that has nothing to do with entropy.

besides these few expectation values about the firms in the industry. So in particular, we know nothing about the size of the total labor pool, internal operating details of the firms, etc. For simplicity, we are considering the case where our ignorance is so broad that we do not even know any stylized facts about the kinds of principle-agent problems holding for firms, anything about how quickly they can change the composition of their workforce as they interact with one another, etc.

How should we, external to the game and given only these expectation values, predict $q(n) = \prod_i q_i(n_i)$, the probability distribution of firm i having n_i^j employees of type j? As before we can do this with the Maxent procedure we used above in deriving the QRE. In the development here n is the state space variable rather than x, which was the state space variable in the derivation of the QRE above. Similarly $u_i(n)$ plays the same role here that $u_i(x)$ does in the development above, and the vector n_i plays the role of the pure strategy x_i above. So writing it out in full, the Maxent Lagrangian for firm i is

$$\mathscr{L}_i(q_i) = S(q_i) + \beta_i\Big[K_i - \sum_{n_i} \mathbb{E}(u_i \mid n_i) q_i(n_i)\Big] + \mu_i\beta_i\Big[N_i - \sum_{n_i}\Big(q_i(n_i)\sum_j n_i^j\Big)\Big] + \lambda'_i\Big[1 - \sum_{n_i} q_i(n_i)\Big] \quad (3)$$

where

$$\mathbb{E}(u_i \mid n_i) = \sum_{n_{-i}} u_i(n_i, n_{-i}) q_{-i}(n_{-i}), \quad (4)$$

the values λ'_i, β_i and μ_i give us the Lagrange parameters, and as in statistical physics we have adopted the convention of writing the Lagrange parameter for the constraint on expected counts as the product of two of those values.

Evaluating the associated Lagrange equations results in a "grand canonical" QRE, given by the following set of F coupled nonlinear equations:

$$q_i(n_i) = \frac{e^{\beta_i \mathbb{E}(u_i|n_i) + \beta_i\mu_i \sum_j n_i^j}}{\sum_{n'_i} e^{\beta_i \mathbb{E}(u_i|n'_i) + \beta_i\mu_i \sum_j n'^j_i}} \quad (5)$$

where the values $\{\beta_i, \mu_i : i = 1, \ldots, F\}$ are set by the provided expectation values and the provided expected total number of employees of each firm. The coupling arises through the term $\mathbb{E}(u_i \mid n_i^j)$, since it depends on the vectors $q_k(n_k)$ for $k \neq i$ in general. (This is just like how in the QRE, $\mathbb{E}(u_i \mid x_i)$ depends on the values $q_k(x_k)$ for $k \neq i$ in general.)

Note that in deriving this result we do not assume that firm i in any sense "chooses" to have a particular vector of employee numbers n_i; we do not anthropomorphize firms. Rather the distribution q_i solely reflects lack of information of *the scientist external to the game* who is making predictions concerning the behavior of the firms.

In some cases the external scientist will have more information than expected values of utility functions and number of employees (e.g., information about the internal structure of the firms, information about the size of the labor market, information about the values of higher order moments of utility functions beyond first order expectations, etc.). Whenever such extra information can be expressed as inequality constraints involving q, the Maxent procedure for formulating q changes in a straightforward way: one expands the Lagrangian to include those constraints, so that they appear as terms in the exponentials giving the separate q_i.

Future work on Maxent noncooperative equilibria involves incorporating the extensive experimental data concerning human behavior [22] as additional constraints for the Maxent procedure. The resultant Maxent solution could be viewed as a refined version of our behavioral models concerning bounded rationality.

THE SHAPE OF THE QRE SURFACE

In the rest of this paper we concentrate on the conventional QRE with a fixed number of players, rather than consider the grand canonical QRE. To analyze the QRE surface of Eq. 2, we express that equation as a set of functional relationships,

$$f_i[q_{-i}, \beta_i] - q_i = 0 \quad (6)$$

for all players i and associated vectors q_i and q_{-i}. For example, when there are only two players, by choosing either player as i and then plugging in twice we get the equation

$$f_i[f_{-i}(q_i, \beta_{-i}), \beta_i] - q_i = 0. \quad (7)$$

This gives q_i as a function of itself and of the two β's. Implicit differentiation then tells us that the function from (β_i, β_{-i}) to q_i is ill-behaved at any point where

$$\frac{\partial f_i}{\partial q_{-i}} \frac{\partial f_{-i}}{\partial q_i} \frac{\partial q_i}{\partial \beta_i} + \frac{\partial f_i}{\partial \beta_i} - \frac{\partial q_i}{\partial \beta_i} = 0 \quad (8)$$

cannot be solved for $\frac{\partial q_i}{\partial \beta_i}$, i.e., where $\det(\frac{\partial f_i}{\partial q_{-i}} \frac{\partial f_{-i}}{\partial q_i} - \mathrm{Id}) = 0$.

To illustrate this and related phenomena, we consider some games between a Row and Column player where each player has only two pure strategies. The first game we consider is the famous "battle of the sexes" coordination game [5]. In this game the utility functions of the players can be represented as

$$\begin{matrix} 2|1 & 0|0 \\ 0|0 & 1|2 \end{matrix} \quad (9)$$

where the first (second) entry in each cell is the Row (Column) player's utility for the associated pure strategy profile.

The q in Eq. (2) induces a value of $\mathbb{E}(u_i)$ given by

$$\sum_{x_i} \mathbb{E}(u_i \mid x_i) q_i(x_i) = \sum_{x_i} \mathbb{E}(u_i \mid x_i) \frac{e^{\beta_i \mathbb{E}(u_i|x_i)}}{\sum_{x'_i} e^{\beta_i \mathbb{E}(u_i|x'_i)}} \equiv \kappa_i^u(\beta_i). \quad (10)$$

where u is shorthand for the set of utility functions of all the players. $\kappa_i^u(.)$ is a monotonically increasing function.[7]

[7] To see this, note that for any set of utility functions and number of players, $\partial\kappa_i^u(\beta_i)/\partial\beta_i$ equals the variance of the Boltzmann distribution given in Eq. (2).

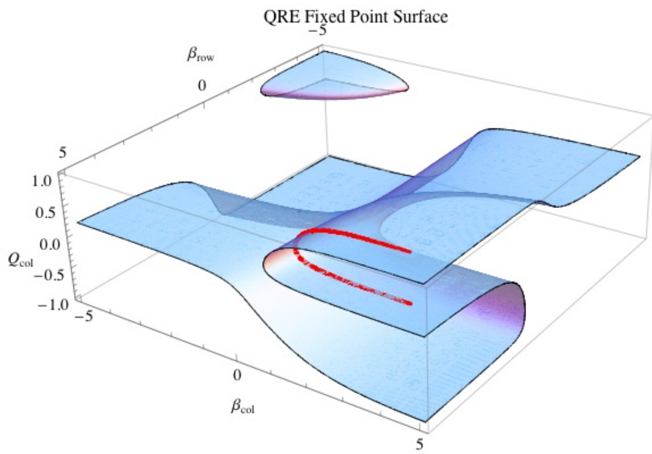

FIG. 1. Col's mixed strategy plotted against $\vec{\beta}$. This is for the QRE of the game in Eq. 9. The hysteresis path discussed in the text is highlighted, starting at the bottom-right.

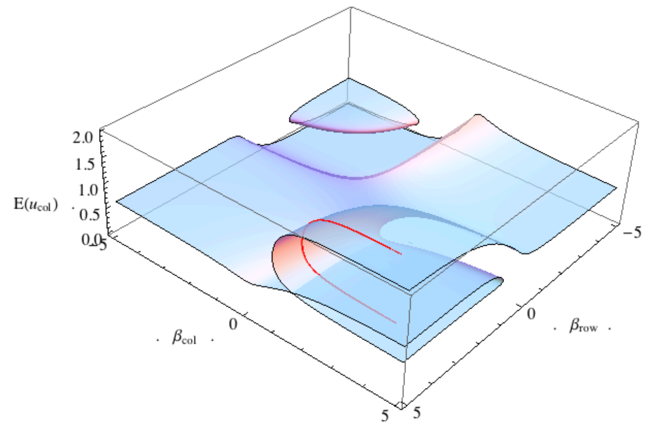

FIG. 2. $\mathbb{E}(u_{col})$ vs. $\vec{\beta}$ under the QRE of the game in Eq. 9. The hysteresis path discussed in the text is highlighted, starting at the bottom-right.

Therefore it is invertible over its codomain, $[\min_{x_i}\mathbb{E}(u_i \mid x_i), \max_{x_i}\mathbb{E}(u_i \mid x_i)]$. So each possible vector of constraints $\vec{K} = (K_{row}, K_{col})$ implicitly sets an associated vector of inverse temperature $\vec{\beta} \equiv (\beta_{row}, \beta_{column})$. This inverse temperature vector in turn fixes the QRE q's for the game.

This means we can consider the functions taking $\vec{\beta}$ to the QRE q's and associated expected utilities, rather than the functions taking $\vec{K}$ to the QRE q's and associated expected utilities. Fig. 1 plots the surface taking $\vec{\beta}$ to Col's mixed strategy, expressed as the symmetrized variable $Q_{Col} \equiv 2[q_{Col}(x_2) - q_{Col}(x_1)]$. Fig. 2 plots the corresponding surface taking $\vec{\beta}$ to $\mathbb{E}_q(u_{col})$.[8]

There are three NE of this game: one where the players jointly follow the pure strategy profile that Row wants (Top-Left), one where they jointly follow the pure strategy profile Col wants (Bottom-Right), and one where there is 2/3 probability of Row choosing Top, and 2/3 probability of Col choosing Right. These three NE correspond to the the three folds of the surface in the bottom right sections of Fig 1 and Fig. 2. The *bottom* fold of Fig. 2 corresponds to the uniform random NE, which is the *middle* fold in Fig. 1. (Note that the battle-of-the-sexes game is not a zero sum game; here the uniform-mixing NE is the worst of the three NE for both players.) Examination of the full surface in Fig. 1 shows that there is no connected path that:

1. is restricted to the quadrant where both $\beta_i > 0$;
2. starts at one of the two pure strategy NE;
3. ends at the other pure strategy NE;
4. only involves changes to one player's β.

Any connected path from one pure strategy NE to the other pure strategy NE involves changes to both β_i.

At bifurcations the number of QRE solutions changes between one and three. This means that infinitesimal changes in $\vec{\beta}$ may result in discontinuous changes in expected utility. As an example, this happens if the system starts at $\vec{\beta} = (5, 5)$ on the top surface, and then β_{row} is reduced to 0.

More generally, if one and/or the other player gradually changes their rationality value β_i, then the system will follow a path on the surfaces. Such paths can be quite complex, depending on the precise trajectory through rationality space. For example, say we start in the region where β_{row} is near 4 and β_{col} is near 4, and that the QRE is on the lowest of the three folds in that region in Fig. 2. Fix β_{row}, and start to decrease β_{col}, as illustrated in the figures. As the column player makes these changes to her rationality, $\mathbb{E}(u_{col})$ gradually increases. By appropriately slowing her changes to her rationality and eventually starting to increase β_{col} again, the column player could cause the path followed by the joint behavior of the players to "round the bend" in the surface. Doing this puts the two players in the top fold of the plot, and as the column player continues increasing her rationality she (still) increases her expected utility.[9]

As an alternative to following the bend though, the column player could monotonically decrease her rationality. Eventually this would cause the two players to fall off the edge of the fold (go through the bifurcation), and fall to the fold that is the bottom of both Fig. 2 and Fig. 1. This will cause the column player to experience a discontinuous fall in her expected

---

[8] Note that the QRE equations are not changed if one interchanges both the utility functions and the β's of the players. Therefore the same plot gives the expected utility of the Row player, if one flips the β axes.

[9] Note that there is no dynamic model underlying our derivation of the QRE surface, so the dynamic stability of each point on the indicated path is undefined. However in many dynamic models the middle fold of Fig 1 — where the indicated path starts — is unstable. In general though, given a dynamic model, there will be paths somewhere in the figure that both exhibit hysteretic effects like that of the indicated path and are stable everywhere.

utility. Moreover, if she were to continue to decrease her rationality after that fall, she would decrease her expected utility further. So after the fall, it would make sense for her to start increasing her rationality, just as when she had followed the bend. Having gone through the discontinuous fall though, in this path the system is on the bottom fold of Fig. 1 (the middle fold of Fig. 2) rather than the top one, to the column player's detriment.

The negative temperatures (negative β's) in the plots correspond to "anti-rational behavior". In such behavior, the associated player is more likely to pick the pure strategy that is worst for them, given what the other player chooses. This may happen, for example, due to social norms.

In this regard, an interesting effect occurs if we multiply the utilities by -1. Fig. 4 illustrates part of the surface after this switch. Note that on the bottom fold, for fixed β_{col}, decreasing β_{row} increases $\mathbb{E}(u_{row})$. So Row benefits by being *less* rational, due to how Column responds to Row's drop in rationality. In essence, it is smart to be dumb, for that player.

MODIFYING GAME PARAMETERS TO IMPROVE SOCIAL WELFARE

Changing parameters of the underlying game

Say an agent external to a game wishes to modify the joint behavior of the two players, preferring some behaviors to others. Suppose as well that the agent can modify the game the agents are playing, e.g., by modifying some parameters of the utility functions of the players. How can the external agent use such ability to change the utility functions to induce the players to change their joint behavior to new behavior that the external agent prefers?

To answer this question we have to model what about player behavior is invariant under the class of changes to their environment that the external agent can impose.[10] More precisely, we have to model what about the Maxent constraints on the players is invariant as the game they are playing changes. In particular, since the underlying utility functions are changing, we cannot require that the expected values of the utilities are invariant; we must specify a different aspect of the constraints to be invariant.

To do this we interpret the constant in the QRE exponent of each q_i as a behavioral attribute of that player which quantifies their "rationality", in the sense of quantifying how close to optimal their mixed strategy is. Under this interpretation, the constant in the QRE q_i's exponent cannot be changed by the external agent. In particular it is independent of changes the external agent can make to the utility functions. (Since each such constant plays the role of an inverse temperature, requiring that they be invariant under changes to the utility functions is analogous to modeling a change to the environment of a thermodynamic system as being isothermal.)

[10] This is analogous to modeling what about a thermodynamic system is invariant under a given class of changes to the system's environment, e.g., modeling whether the changes are adiabatic, isothermal, etc.

To capture this restriction means we must re-express our constraints. There are an infinite number of ways that we might do this. In particular, a natural choice is to require that for any utility function w_i for player i, q_i must obey

$$\mathbb{E}_{q_i}(w_i) = \kappa_i^w(b_i) \tag{11}$$

for a fixed value of i's rationality constant, b_i (where q_{-i} is implicit as usual).

As an illustration, suppose that while they cannot affect behavioral attributes like b_i, the external agent can apply a player-specific "tax rate" $1 - \alpha_i$ to each player i. Formally, this means that the utility function for player i changes from u_i to $v_i \equiv \alpha_i u_i$. This means the Maxent Lagrangian for each player i becomes

$$\mathscr{L}_i(q_i) = S(q_i) + \lambda_i[\mathbb{E}_{q_i}(v_i) - \kappa_i^v(b_i)] + \lambda_i'[1 - \sum_{x_i} q_i(x_i)] \tag{12}$$

giving the solution

$$q_i(x_i) \propto exp[\lambda_i \mathbb{E}(v_i \mid x_i)] \tag{13}$$

where λ_i is set by the constraint $\mathbb{E}(v_i) = \kappa_i^v(b_i)$. The expected utility for this q is

$$\begin{aligned}\mathbb{E}(v_i) &= \frac{\sum_{x_i} \mathbb{E}(v_i \mid x_i)\; exp[\lambda_i \mathbb{E}(v_i \mid x_i)]}{\sum_{x_i} exp[\lambda_i \mathbb{E}(v_i \mid x_i)]} \\ &= \kappa_i^v(\lambda_i).\end{aligned} \tag{14}$$

Since $\kappa_i^v(.)$ is invertible, the only way this is possible in light of our constraint that $\mathbb{E}(v_i) = \kappa_i^v(b_i)$ is if $\lambda_i = b_i$. Accordingly the solution is

$$\begin{aligned}q_i(x_i) &\propto exp[b_i \mathbb{E}(v_i \mid x_i)] \\ &= exp[\alpha_i b_i \mathbb{E}(u_i \mid x_i)]\end{aligned} \tag{15}$$

with b_i a fixed attribute of player i.

Comparing Eq.'s (2) and (15), we see that changing α_i while leaving b_i fixed, for the version of a game that has variable tax rates $1 - \alpha_i$, is the same as changing β_i, for the version of that game that has taxes fixed to zero. (Intuitively, changing the tax on a player is the same as changing how rational they are.) So in particular if the untaxed version of the game is the one specified in Eq. 9, then the surface plotting QRE values of $\mathbb{E}(v_i)$ as a function of (α_i, α_{-i}) in the taxed version of the game is given by a simple transformation to the surface in Fig. 2. The resultant new surface is shown in Fig. 3, for the choice that $b_i = 5$ for both players.

From now on, for pedagogical simplicity, rather than distinguish α_i and b_i we will simply work with their product. We will label this product as "β_i". Also for simplicity we will make references to the figures which have β_i as independent variables even when we are concerned with varying α_i. When we do this the multiplication of utility values by α_i — which

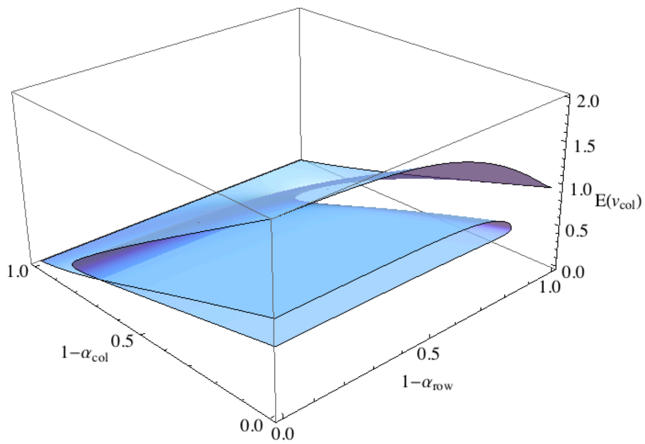

FIG. 3. Col's expected post-tax utility plotted against the tax rates (expressed as fractions) on the two players. This is for the QRE of the game in Eq. 9. The constraint for each player i in this taxed game is $b_i = 5$. So for example, $\vec{\alpha} = (1, 1)$ corresponds to the point with $\vec{\beta} = (5, 5)$ in Fig. 2.

amounts to scaling them at each point by the associated value β_i — is implicit.

The first thing to note after recasting the analysis in terms of tax rates is that in general a player i may benefit if her tax rate *increases*. Intuitively, this is because the other player knows that i's tax rate has risen; and therefore makes different predictions for i's behavior; and therefore acts differently herself; all in a way that benefits i more than i loses due to her higher tax rate. An example of this is shown in Fig. 4, where anywhere on the bottom surface, Row benefits if her tax rate increases. (Note that the effect of lowering the parameter β_{Row} on the QRE q of an untaxed game is equivalent to the effect of lowering α_{row} on the QRE q of a taxed version of the game.)

In economic analysis of optimal regulatory policy, typically pairs of exogenous factors like tax rates are compared by examining the joint behavior of the players under the (games parameterized by) those tax rates. This is called "comparative statics" (in contrast to comparative dynamics). The premise is that a regulator should adopt whichever of the exogenous factors results in a higher expected value of some real-valued **social welfare** function defined over that joint behavior [59].

Inspired by comparative statics and the fact that Row may prefer a higher tax rate, we may wonder whether by varying tax rates slowly enough that the joint behavior of the players is always on the QRE surface, we may be able to monotonically improve expected utilities for *both* players. The answer is yes: For some games, by changing tax rates we can gradually move the equilibrium across the surface from one fold to the other, and then undo those changes, returning the rates to their original values, but leaving both players with higher expected utility. (See [46] for other work that exploits the shape of a QRE surface to optimize player joint behavior.)

As an example of this, consider a variant of the battle of the sexes game, where the utility functions are the negatives of the one considered above. The associated bifurcation surface for mixed strategies is the same as the one in Fig. 1, with all β_i multiplied by -1. Similarly, the bifurcation surface for expected utility is the same as in Fig 2, with all β_i multiplied by -1 and the dependent variable of $\mathbb{E}(u_{col})$ also multiplied by -1. For this modified game, there are paths of $\vec{\beta}$'s (i.e, of $\vec{\alpha}$'s) such that:

1. Neither player ever is more rational (taxed at a higher rate) on the path than at the starting point.
2. At each step on the path, if after the next infinitesimal change in $\vec{\beta}$ there is a QRE q infinitesimally close to the current one, it is adopted. (Adiabaticity.)
3. Each infinitesimal change in $\vec{\beta}$ increases both $\mathbb{E}_q(u_i)$'s.
4. At each infinitesimal step, if multiple changes in q meet (1)-(3), but one is Pareto superior to the others (i.e., better for both players), the players coordinate on that one.

Examples of such paths are illustrated in Fig. 4.

Concretely, such adiabatically slow changing of tax rates might be implemented with a large population of players who repeatedly play the game with other anonymous members of the population chosen at random. (Since the players are anonymous, the likelihood of "trigger strategies" or similar phenomena [5–7] that can arise in multi-stage extensive form games should be minimized; the players are likely to treat each game afresh, rather than consider them as stages in such a multi-stage game.) In a first stage of the experiment one would observe the player behavior and use that to statistically estimate their individual rationality coefficients β_i. (See [33, 46] and references therein for how to do such estimation.) One would then change the tax rate a very small amount once every T plays of the game for some T that is large compared to the discounting rate of the players. This would help ensure that the players do not anticipate the future when making their decisions at any particular time.

How best to myopically control a society

The existence of paths through tax space that benefit all players raises the question of how a society should dynamically update its tax rates. We now compare three procedures for how this could be done by society as a whole. (For notational simplicity, and to emphasize the analogy with annealing, we parameterize the procedures in terms of changes to $\vec{\beta}$ rather than changes to $\vec{1} - \vec{\alpha}$.)

I. "Anarchy": Players independently decide how to modify their β's. To do this they follow gradient ascent with a small step size Δ, subject to the constraint that no player i can go to a β_i larger than the starting one. Thus, both players i change β_i by $\delta\beta_i \in [-\Delta, \Delta]$, using $\partial\mathbb{E}(u_i)/\partial\beta_i$ to make their choice of what value $\delta\beta_i$ to pick. (Since this is a linear procedure, the players will always choose one of the three values $\{-\Delta, 0, \Delta\}$.)

II. "Socialism": An external regulator determines the path, again using gradient descent, this time over the sum of the players' expected utilities. At each step of the path $\vec{\beta}$ is changed by the $(\delta\beta_{row}, \delta\beta_{col})$ vector that maximizes

$$[\delta\beta_{row}\frac{\partial\mathbb{E}(u_{row})}{\partial\beta_{row}} + \delta\beta_{col}\frac{\partial\mathbb{E}(u_{row})}{\partial\beta_{col}}] + [\delta\beta_{row}\frac{\partial\mathbb{E}(u_{col})}{\partial\beta_{row}} + \delta\beta_{col}\frac{\partial\mathbb{E}(u_{col})}{\partial\beta_{col}}] \quad (16)$$

subject to $||(\delta\beta_{row}, \delta\beta_{col})||^2 \leq 2\Delta^2$. (The constraint is to match the step size to that of the first procedure.)

III. "Market": Let T be the set of joint expected utilities for all the bargains that a set of N bargainers might reach in a particular bargaining scenario. Then certain mild axioms concerning bargaining behavior of humans give a unique prediction for which bargain in T is reached. This prediction, known as the "Nash bargaining concept" [6, 7], says that the the joint expected utility of the bargain reached is $\text{argmax}_{\vec{u}\in T}[\prod_{i=1}^{N} u_i]$.

We can use the Nash bargaining concept to predict what change to $\vec{\beta}$ the players would agree to under a "market" where they bargain with one another to determine that change. To do this we fix the set of all allowed bargains to the set of all pairs $\vec{\beta}$ such that $||\vec{\beta}-\vec{\beta}(t)||^2 \leq 2\Delta^2$, where $\vec{\beta}(t)$ is the current joint β. We also choose $\vec{d}$ to be the joint expected utility at $\vec{\beta}(t)$. So under Nash bargaining, at each iteration t, the players choose the change in joint β, $\delta\vec{\beta}$, that maximizes the product

$$\left[\mathbb{E}(u_{Row} \mid \vec{\beta}(t) + \delta\vec{\beta}) - \mathbb{E}(u_{Row} \mid \vec{\beta}(t))\right] \times \left[\mathbb{E}(u_{Col} \mid \vec{\beta}(t) + \delta\vec{\beta}) - \mathbb{E}(u_{Col} \mid \vec{\beta}(t))\right] \quad (17)$$

subject to $||\delta\vec{\beta}|| \leq 2\Delta^2$.

As in the other two procedures, we use first order approximations in this one, to evaluate the two differences in expected utilities. This means that in this procedure, we find the vector $\delta\vec{\beta}$ that maximizes the product

$$\left[\nabla_{\vec{\beta}}\mathbb{E}(u_{Row} \mid \vec{\beta}) \cdot \delta\vec{\beta}\right]\left[\nabla_{\vec{\beta}}\mathbb{E}(u_{Col} \mid \vec{\beta}) \cdot \delta\vec{\beta}\right] = \left[\delta\beta_{row}\frac{\partial\mathbb{E}(u_{row})}{\partial\beta_{row}} + \delta\beta_{col}\frac{\partial\mathbb{E}(u_{row})}{\partial\beta_{col}}\right] \times \left[\delta\beta_{row}\frac{\partial\mathbb{E}(u_{col})}{\partial\beta_{row}} + \delta\beta_{col}\frac{\partial\mathbb{E}(u_{col})}{\partial\beta_{col}}\right] \quad (18)$$

subject to $||(\delta\beta_{row}, \delta\beta_{col})||^2 < 2\Delta^2$. Using Lagrange multipliers, for given $\nabla_{\vec{\beta}}\mathbb{E}(u_{Row} \mid \vec{\beta})$ and $\nabla_{\vec{\beta}}\mathbb{E}(u_{Col} \mid \vec{\beta})$, the optimal $\delta\vec{\beta}$ is given by solving a quadratic equation. Once we have found that optimal $\delta\vec{\beta}$, we add it to $\vec{\beta}(t)$ to get the new position in β space. (Compare Eq. (18) to Eq. (16).)

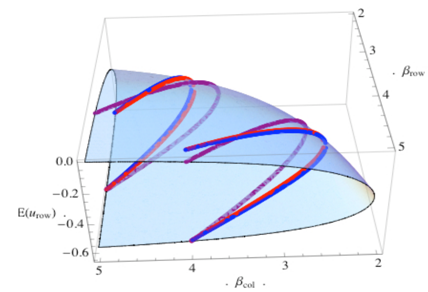

FIG. 4. A QRE surface with paths shown for the anarchy (red), socialism (blue) and market (purple) procedures. As in Fig. 2, the x and y axes are player rationalities, β_{row} and β_{col}, and the z axis is expected utility (this time of player Row). This plot is for a variant of the Battle of the Sexes game where all utility functions are multiplied by -1.

A variant of this market procedure could be used as to model how democracies would modify the game parameters, or more generally to model any process by which the players of the game collectively gradually change the parameters of the game.

Note that in all three procedures the total change in $\vec{\beta}$ in any step never exceeds $\sqrt{2}\Delta$. This adiabaticity reduces the computational burden on the players, by not changing the game too much from one timestep to the next.

As in standard economics, we can quantify how good a full path produced by a procedure is for society as a whole by calculating the discounted sum of future social welfare along the path, e.g., by defining social welfare as the sum of player utilities:

$$W \equiv \sum_{t'>0}(1+\gamma)^{t-t'}\sum_{i=1}^{N}\mathbb{E}(u_i(t')) \quad (19)$$

Using this definition we can compare the three procedures by calculating the W's for the paths they generate starting from some shared $\vec{\beta}$ at time $t = 0$. We did this for two representative initial $\vec{\beta}$'s, for the surface in Fig. 4, with the resultant paths illustrated in that figure.

While for both of the initial $\vec{\beta}$'s any two of the paths will intersect at some point $(\vec{\beta}, q)$, they get to that intersection point at different times. In addition, they diverge beyond that intersection point. These two effects mean that the discounted sum of future expected utilities is different for the three procedures of changing $\vec{\beta}$.

We found that anarchy always did worse than the other two procedures. Those others are compared to each other in Fig. 5. When the discounting factor γ is large (i.e., we are more concerned with near-term than long-term utility) the market procedure does better, otherwise socialism does.

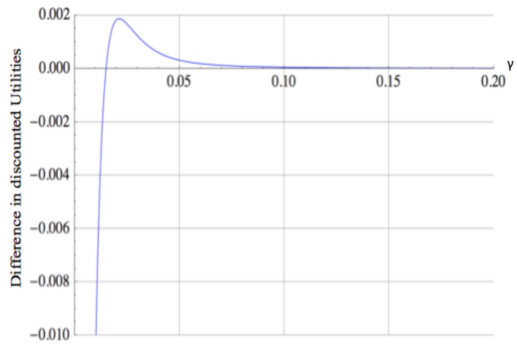

FIG. 5. The difference between the discounted sums of future expected utilities of the two players under the "socialism" and "market" procedures, plotted against the discounting factor γ.

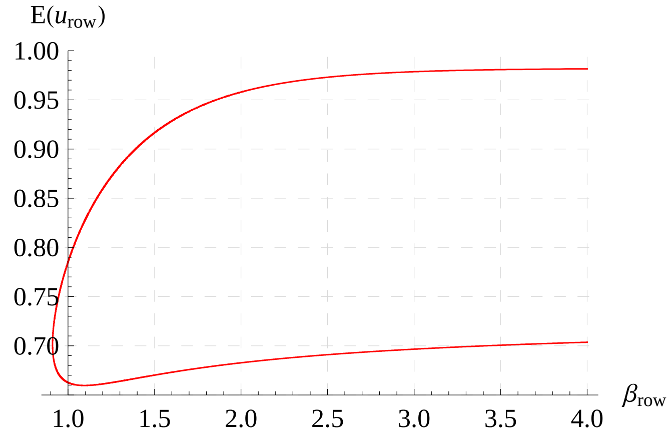

FIG. 6. The expected utility of the Row player along the path through $\vec{\beta}$ highlighted in Fig. 2, illustrated as a function of the Row player's rationality, β_{row}. The path starts at the bottom right, then travels left, before turning and finishing at the top right.

To our knowledge no laboratory game theory experiment has ever looked at slowly varying game parameters [22, 60] (with the partial exception of [57], which as mentioned above was posted subsequent to the posting of this paper). However it would not be difficult to design such an experiment. For example, we could have some of the experiments be where the tax rates are changed in an "anarchic" manner, where the players can change their tax rates a very small amount at some regular interval, but are not allowed to interact to decide those tax rates; some where the players are allowed to bargain on those changes; and some where the changes are set by an external "socialist" regulator, based on their QRE model of player behavior.[11]

Of course, even if such experiments were to confirm the predictions made by our theoretical analysis, the results should in no way be taken to imply how society should change tax rates; the toy models considered in this paper are illustrative only. An interesting question in this regard is what characteristics of a (potentially very realistic) game determine the relative performances of the three control algorithms we have considered. Even more interesting would be to investigate this issue for more realistic control algorithms than the ones we consider.

[11] As a practical matter, we may have to suggest ahead of time to the subjects that it may be in their interest to have their own tax rate increased, since that is counter-intuitive. Also, to make sure the games start at a non-optimal QRE for the initial tax rates, we may want to start each subject playing against a computer programmed to behave as the subject's opponent would at that non-optimal equilibrium. Then as the experiment progresses, we could simultaneously replace the programs playing against some pair of subjects A and B with those subjects themselves, i.e., suddenly have the subjects play one another rather than computer programs.

How best to non-myopically control a society

All three procedures described above are local, only looking a single step into the future, and therefore only considering the local shape of the QRE surface. A procedure that also considers the QRE surface's global geometry will produce better paths in general. In particular, such global information allows us to consider paths where a player loses expected utility for certain periods, but in the end all players are better off. Fig. 2 highlights such a path, along which player Column always benefits but player Row loses initially, before ultimately benefiting. A cross-section of the expected utility of Row along the path is shown in Fig. 6. Note that player Row might demand compensation to agree to follow such a path where they temporarily lose expected utility, e.g. in terms of a subsidy paid for with a bond that is repaid by all players at the end of the path.

Once we allow such paths whose benefits arise from the global rather than local geometry of the QRE surface, we are faced with the question of which path should be adopted for any given starting point of the path. Under a socialism model, this question is relatively well-posed. For example, we could stipulate that the path followed be the one that maximizes the discounted sum of utilities, either with subsidies taken into account or not.

In general, by implementing the associated sequence of regulations the regulator would induce higher social welfare than they could induce if they instead used a comparative statics approach, in which they implement a single change in the regulations. Indeed, it is not even clear how one could use a single regulation change to do some of the things possible with an extended path through regulation space. For example, say the regulator is told to get society from joint behavior lying on a suboptimal fold for a current value of a regulation parameter to joint behavior lying on the optimal fold *for that same value* *of that regulation parameter*. One can often do this by following a path through regulation space, as illustrated in Fig. 2 for the case where the regulations are tax rates. However one cannot do this with a single change in regulations.

There should also be practical advantages to implementing a sequence of small changes to regulations rather than a single large change to regulations. In particular, doing so would allow the regulator to modify their behavioral model of the players as the sequence unfolds, and thereby improve that sequence. For example, they could refine their estimates of exponents β_i as the sequence unfolds, and therefore refine their estimate of the QRE surface. They would then use that improved estimate of the QRE surface to improve the remaining regulation changes in the sequence. Alternatively, as the sequence unfolds they might acquire extra information for the Maxent procedure beyond the constraints concerning expected utilities of the players. This too would cause them to change their estimate of the surface taking regulations to expected utilities of the players, and therefore cause them to modify the remaining steps in the sequence of regulation changes. (In fact, when they acquire new constraints for the Maxent procedure, the surface they use would no longer be a QRE surface.)

In these kinds of ways, the regulator could exploit feedback when using a sequence of small changes in regulations to control joint behavior of the players. In contrast, in an approach involving a single change in regulations, the control is purely "open-loop". (Some of these advantages of "gradualist" regulatory policies that implement paths through regulation space have been discussed in the economics references mentioned above, albeit under assumptions of fully rational behavior in "coarse-grained", macroeconomic models.)

While the issues in using global properties of the QRE surface to determine paths through regulation space are relatively well-posed for the socialism model of regulation, under the market model they become more open-ended. That is because in a market model, all players have to agree to the path. So rather than a sequence of bargains, each over an infinitesimal step along the path (as in the analysis above), the players would bargain over the entire path at once. For example, such bargaining might be modeled by saying that each player values any given full path as the discounted sum of their future utilities along that path. Under this model, the joint valuation of any given path is given by the vector of all players' future-discounted sums of utilities for that path. The **feasible set** of possible joint valuations that underlies the bargaining is the set of possible joint valuations for all paths. At any given moment, the players would bargain over which element in the feasible set to adopt. One could then predict what bargain they reach using the Nash bargaining solution, for example.

Note though that this model opens the issue of "hold-up" problems, where once a path has been followed a certain distance, the relative bargaining powers of the players for the remainder of the path changes. More precisely, say that at $t = 0$ there is a joint rationality $\vec{\beta}(0)$, and that society starts to follow a path $\vec{\beta}_0(t)$ from there that is a Nash bargaining solution at $t = 0$ over the feasible set $T(0)$ given by all paths starting at $\vec{\beta}(0)$. Then in general, for $t' > 0$, the path $\vec{\beta}_{t'}(t)$ that is a Nash bargaining solution for full paths starting from $\vec{\beta}_0(t')$ is not a truncation of $\vec{\beta}_0(t)$ to $t > t'$. (This is because the feasible set $T(t)$ of possible remaining joint values of each player's future-discounted sums of utilities may change its shape as the path is traversed, not just get rescaled.) There is an inconsistency across time.

The analysis becomes even more complicated if the players intermix their bargaining over the parameters of the game with their strategies in that game. For example, it may be that in choosing their mixed strategy at any time step t, a player i would consider how the ensuing feasible set $T(t + \delta t)$ depends on their choice. More generally, if they are bargaining over full paths, the players may be forced to consider entire sequences of strategies in the associated set of games, rather than treat each successive game independently. (This means that their "strategy space" is far more complicated than in the simple case analyzed above where players only use local properties of the QRE surface.) All of this suggests the analysis should include game theoretic concepts like binding commitments, renegotiation-proof equilibria, etc. The situation gets even richer if paths involving subsidies are allowed. All of this is the subject of future work.

FINAL COMMENTS

In this paper we have focused on how the Maxent procedure of statistical physics can be applied to noncooperative game theory. The importance of Jaynes use of Maxent to derive the canonical ensemble distribution is that it provided a new perspective on that distribution, as arising from the fundamentally statistical nature of the scientists task rather than intrinsic fluctuations of the system. This allowed him to sidestep various controversial presumptions in earlier derivations, e.g., involving ergodicity or heat baths (which for example are nonsensical when the system in question is the entire universe).

Similarly, our derivation provides a new perspective on the QRE distribution, as arising from the fundamentally statistical nature of the scientist's task rather than intrinsic fluctuations of they system. This allows us to sidestep various controversial presumptions in earlier derivations, e.g., that the players are completely rational, or have played the same game with one another an infinite number of times.

We then showed that the QRE distribution contains some phenomena quite familiar from statistical physics, like bifurcation surfaces and hysteresis effects. However in some ways it is intrinsically more complicated than conventional statistical physics, since it involves multiple utility functions rather than a single Hamiltonian. We then went on to explore some toy models of the implications of this Maxent game theory formalism for issues of how to manage a society.

There are many other issues one could investigate with this formalism however. To give a simple example, in the real world, whatever process might change game parameters, it might be quite noisy. The QRE surface provides information about how stable player behavior should be against such noise in the game parameters. For example, say the players are on the top fold of the surface in Fig. 2, with $\vec{\beta} = (2, 4)$, so the joint behavior is near an edge of the QRE surface. In this situation, small external noise may lead the players to "fall off the edge", and undergo a discontinuous jump to the lower surface. Moreover, even if the players managed to (adiabatically slowly) restore their original rationalities after such a jump, they would end up on the middle fold of the region where β_{row} is near 2, not on the good fold they started in. Due to this, when an economic situation exhibits such qualitative features, it may behoove society to stay away from such edges in the QRE surface, even if that lowers total expected utility.

Acknowledgments: *This work was carried out under the auspices of the National Nuclear Security Administration of the U.S. Department of Energy at Los Alamos National Laboratory under Contract No. DE-AC52-06NA25396. We would like to thank John Miller and Sam Bowles for helpful comments.*

* Center for Nonlinear Studies, MS B258, LANL, Los Alamos, NM, 87545; NASA Ames Research Center, MailStop 269-1, Moffett Field, CA 94035-1000, `david.h.wolpert@nasa.gov`

† Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA; Department of Mathematics and Computer Science, Leipzig University, 04009 Leipzig, Germany

‡ Dept. of Physiology, Sydney Medical School, The University of Sydney, Australia

[1] T. Cover and J. Thomas, *Elements of Information Theory* (Wiley-Interscience, New York, 1991).

[2] D. Mackay, *Information Theory, Inference, and Learning Algorithms* (Cambridge University Press, 2003).

[3] E. T. Jaynes and G. L. Bretthorst, *Probability Theory : The Logic of Science* (Cambridge University Press, 2003).

[4] E. T. Jaynes, Physical Review **106**, 620 (1957).

[5] D. Fudenberg and J. Tirole, *Game Theory* (MIT Press, Cambridge, MA, 1991).

[6] R. B. Myerson, *Game theory: Analysis of Conflict* (Harvard University Press, 1991).

[7] M. Osborne and A. Rubinstein, *A Course in Game Theory* (MIT Press, Cambridge, MA, 1994).

[8] R. Aumann and S. Hart, *Handbook of Game Theory with Economic Applications* (North-Holland Press, 1992).

[9] J. Hofbauer and K. Sigmund, *Evolutionary games and population dynamics* (Cambridge Univ.Press, 1998).

[10] J. Maynard Smith, *Evolution and the theory of games* (Cambridge Univ.Press, 1982).

[11] D. Foley, Journal of Economic Theory **62**, 321 (1994).

[12] M. Ausloos, "Econophysics of stock and foreign currency exchange markets," in *Econophysics and Sociophysics* (Wiley-VCH Verlag GmbH & Co. KGaA, 2006) Chap. 9, pp. 249–278.

[13] E. Samanidou, E. Zschischang, D. Stauffer, and T. Lux, Reports on Progress in Physics **70**, 409 (2007).

[14] G. Weisbuch, "Social opinion dynamics," in *Econophysics and Sociophysics* (Wiley-VCH Verlag GmbH & Co. KGaA, 2006) Chap. 12, pp. 339–366.

[15] D. Sornette, *Why stock markets chrash: Critical Events in Complex Financial Systems* (Princeton University Press, 2004).

[16] D. Challet, Journal of Economic Dynamics and Control **32**, 85 (2008).

[17] S. Alfarano, T. Lux, and F. Wagner, Journal of Economic Dynamics and Control **32**, 101 (2008).

[18] M. Levy, Journal of Economic Dynamics and Control **32**, 137 (2008).

[19] J. Farmer, M. Shubik, and D. E. Smith, "Economics: The next physical science?" (2005), sFI working paper 05-06-027.

[20] D. H. Wolpert, in *Complex Engineered Systems: Science meets technology*, edited by D. Braha, A. Minai, and Y. Bar-Yam (Springer, 2004) pp. 262–290.

[21] C. Shannon and W. Weaver, *The mathematical theory of communication* (Univ. Illinois Press, 1949).

[22] C. Camerer, *Behavioral Game Theory: Experiments in Strategic Interaction* (Princeton University Press, 2003).

[23] C. Starmer, Journal of Economic Literature **38**, 332 (2000).

[24] J. von Neuman and O. Morgenstern, *Theory of Games and Economics Behavior* (Princeton university Press, 1944).

[25] C. Aliprantis and K. C. Border, *Infinite Dimensional Analysis* (Springer Verlag, 2006).

[26] D. Fudenberg and D. K. Levine, *The Theory of Learning in Games* (MIT Press, Cambridge, MA, 1998).

[27] S. Hart, Econometrica **73**, 1401 (2005).

[28] A. Rubinstein, *Modeling Bounded Rationality* (MIT press, 1998).

[29] S. Russell and D. Subramanian, Journal of AI Research , 575 (1995).

[30] M. Georgeff, B. Pell, M. Pollack, M. Tambe, and M. Wooldridge, in *Intelligent Agents V*, Lecture Notes in Computer Science, Vol. 1555 (Springer, Berlin / Heidelberg, 1999) pp. 1–10.

[31] R. D. McKelvey and T. R. Palfrey, Games and Economic Behavior **10**, 6 (1995).

[32] R. D. McKelvey and T. R. Palfrey, Japanese Economic Review **47**, 186 (1996).

[33] R. D. McKelvey and T. R. Palfrey, in *Handbook of Experimental Economics Results*, Vol. 1 (North Holland, 2008) pp. 541–548.

[34] S. P. Anderson, J. Goeree, and C. A. Holt, Southern Economic Journal **69**, 21 (2002).

[35] J. Shamma and G. Arslan, IEEE Trans. on Automatic Control **50**, 312 (2004).

[36] D. Fudenberg and D. Kreps, Games and Economic Behavior **5**, 320 (1993).

[37] J. R. Meginniss, Proceedings of the American Statistical Association, Business and Economics Statistics Section , 471 (1976).

[38] S. P. Anderson, J. Goeree, and C. A. Holt, Scandanavian journal of economics , 21 (2004), preprint title was "Stochastic game theory: adjustment to equilibrum under noisy directional learning".

[39] S. Durlauf, Proceedings Natl. Acad. Sci. USA **96**, 10582 (1999).

[40] R. H. Crites and A. G. Barto, in *Advances in Neural Information Processing Systems - 8*, edited by D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo (MIT Press, 1996) pp. 1017–1023.

[41] J. Hu and M. P. Wellman, in *Proceedings of the Fifteenth International Conference on Machine Learning* (1998) pp. 242–250.

[42] D. H. Wolpert and K. Tumer, Journal of Artificial Intelligence Research **16**, 359 (2002).

[43] D. H. Wolpert, K. Tumer, and E. Bandari, Physical Review E

69 (2004).
[44] D. H. Wolpert, in *Proceedings of WEHIA04*, edited by A. Namatame (Springer Verlag, 2004).
[45] K. E. Train, *Discrete Choice Methods with Simulation* (Cambridge University Press, 2003).
[46] D. Wolpert and N. Kulkarni, in *Proceedings of the 2008 NASA/ESA Conference on Adaptive Hardware and Systems*, edited by A. Erdogan (IEEE Press, 2008).
[47] M. Gocke, Journal of Economic Surveys **16**, 167 (2002).
[48] J. B. Rosser, Journal of Economic Dynamics and control **31**, 3255 (2007).
[49] L. McKenzie, Econometrica **44**, 841 (1976).
[50] T. Bewley, Journal of Mathematical Economics **10**, 233 (1982).
[51] C. Feinstein and S. Oren, Journal of Economic Dynamics and Control , 25 (1985).
[52] W. Brainard, American Economic Review , 411 (1967).
[53] V. H. Dehejia, Economics and Politics **15**, 33 (2003).
[54] V. Popov, Comparitive Economic Studies **39**, 1 (2007).
[55] B. Rosser, M. Rosser, S. Guastello, and R. Bond, Nonlinear dynamics, psychology and life sciences **5**, 345 (2011).
[56] M. Altman, Journal of Socioeconomics **29**, 127 (2000).
[57] J. Romero, "The effect of hysteresis on equilibrium selection in coordination games," (2011), krannert School of Management, Purdue University, paper 1265.
[58] R. B. Myerson, International Journal of Game theory **27**, 375 (1998).
[59] T. J. Kehoe, in *The New Palgrave: A Dictionary of Economics*, edited by J. Eatwell, M. Milgate, and P. Newman (Palgrave Macmillan, 1987) pp. 517–520.
[60] C. Plott and V. Smith, *The Handbook of Experimental Economics Results* (Elsevier Science, 2005).